



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Efficiently serving HDF5 via OPeNDAP

Kent Yang

The HDF Group

This work was supported by NASA/GSFC under
Raytheon Co. contract number NNG15HZ39C

Why OPeNDAP¹?

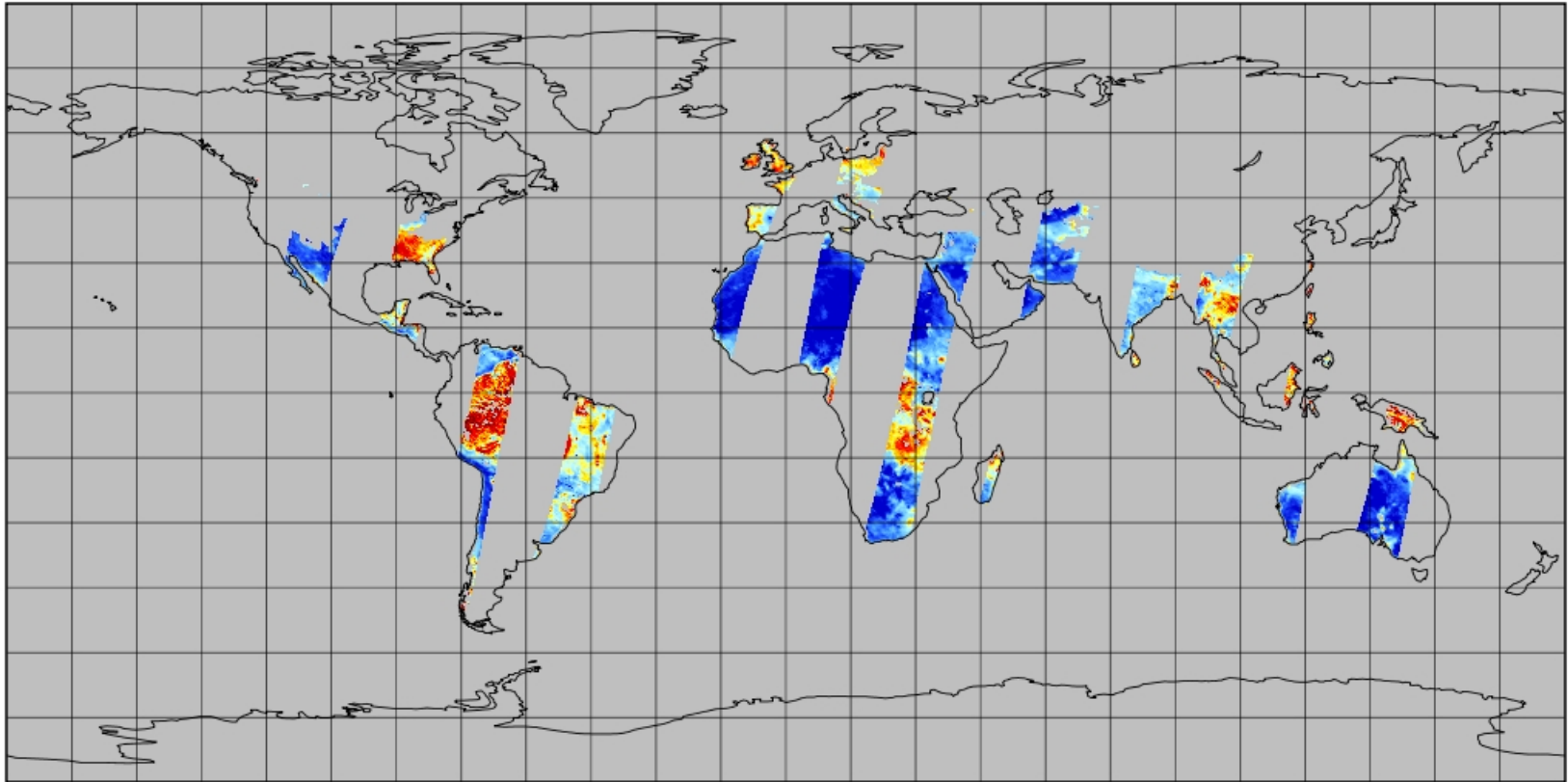
- Check metadata remotely (in various forms)
- Obtain the subset of data easily and efficiently
- Hide the original data sources
 - Hierarchical Data Format (HDF) versions 4 and 5
 - Network Common Data Form (NetCDF)
 - Geospatial Tagged Image File Format (GeoTIFF)
- Many popular earth science tools can visualize and analyze the data via OPeNDAP
- OPeNDAP output (including subsets) can be downloaded as other formats

HDF(5) Hyrax modules

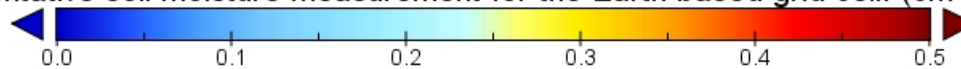
- The visualization of HDF(5) data via Hyrax
 - What NASA users request the most

Visualize A **Soil Moisture Active Passive** (SMAP) HDF5 variable via Hyrax

Representative soil moisture measurement for the Earth based grid cell.



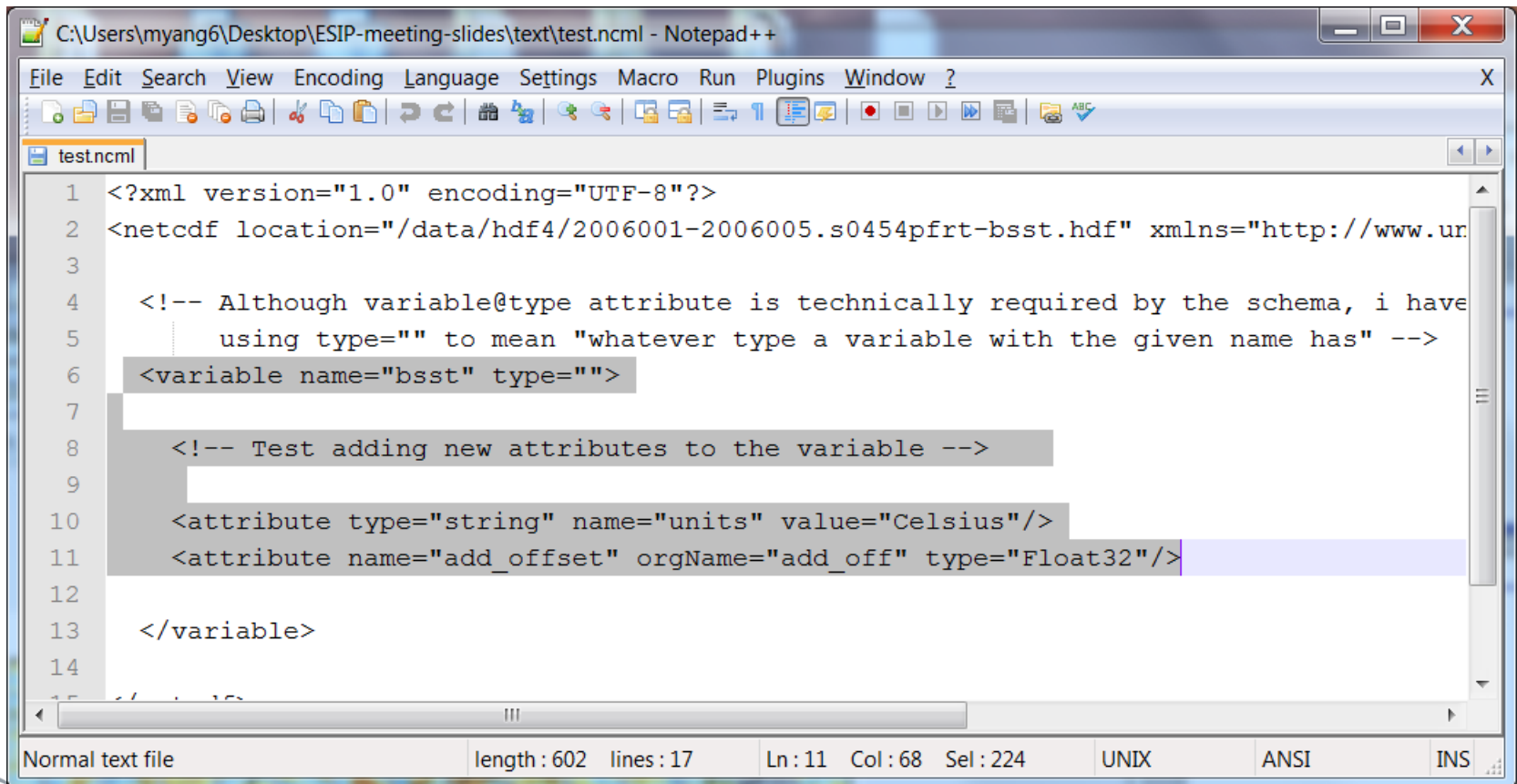
Representative soil moisture measurement for the Earth based grid cell. (cm**3/cm**3)



Data Min = 0.0, Max = 0.5, Mean = 0.2

HDF5 handler and NcML¹

- NcML module can be used with HDF handlers to provide the missing Climate and Forecast Metadata (CF) conventions information



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <netcdf location="/data/hdf4/2006001-2006005.s0454pfrt-bsst.hdf" xmlns="http://www.un
3
4 <!-- Although variable@type attribute is technically required by the schema, i have
5     using type="" to mean "whatever type a variable with the given name has" -->
6 <variable name="bsst" type="">
7
8 <!-- Test adding new attributes to the variable -->
9
10 <attribute type="string" name="units" value="Celsius"/>
11 <attribute name="add_offset" orgName="add_off" type="Float32"/>
12
13 </variable>
14
15
```

Normal text file length: 602 lines: 17 Ln: 11 Col: 68 Sel: 224 UNIX ANSI INS

HDF5 handler and file NetCDF

- File NetCDF module can work with HDF handlers to convert HDF files to NetCDF-3 or NetCDF-4 classic files that follow the CF conventions.

HDF5 handler and file NetCDF

OPeNDAP Server Dataset Query Form - Mozilla Firefox

File Edit View History Bookmarks Tools Help

OPeNDAP Server Dataset Query...

localhost:8080/opendap/data/hdf4/2006001-2006005.s0454pfrt-bsst.hdf.html

OPeNDAP Server Dataset Access Form

Action:

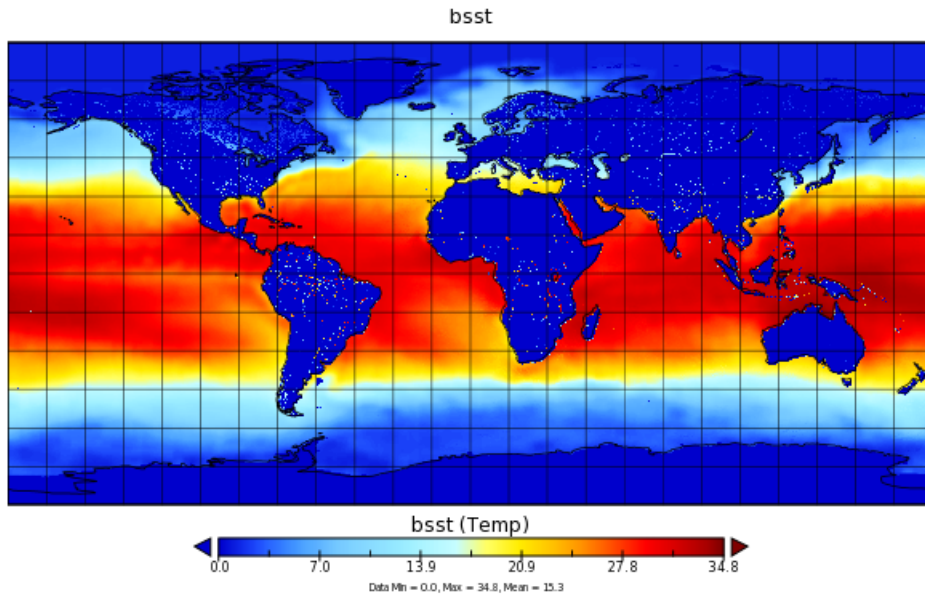
Data URL:

Global Attributes:

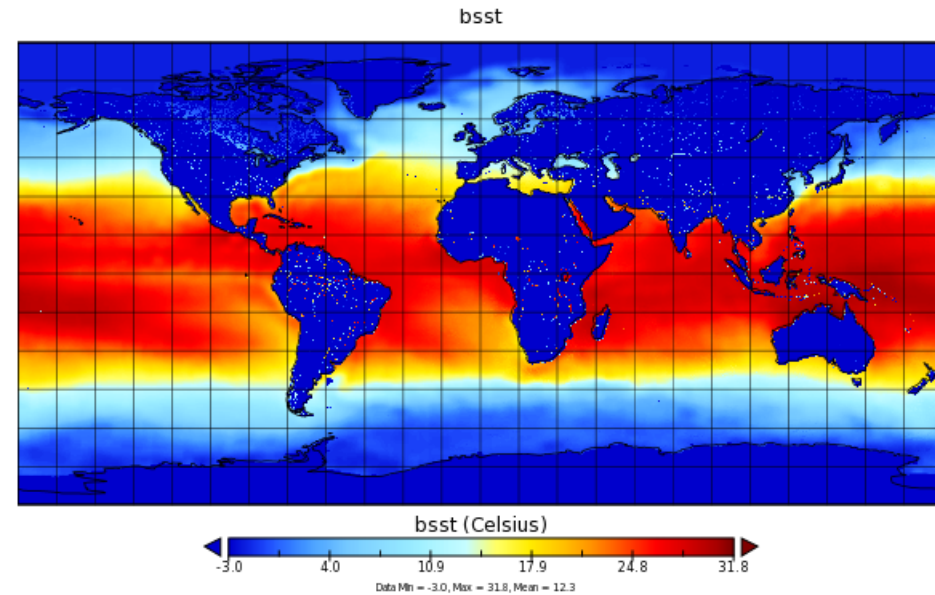
```
HDF_GLOBAL.dsp_SubImageId: 0
HDF_GLOBAL.dsp_SubImageName: BSST
HDF_GLOBAL.dsp_ModificationDate: 20070409
HDF_GLOBAL.dsp_ModificationTime: 120728
HDF_GLOBAL.dsp_SubImageFlag: 64
HDF_GLOBAL.dsp_SubImageTitle: \\\NLC Postage Stamp
```

Variables: ☐ **bsst:** Array of 16 bit Unsigned integers [lat = 0..4095][lon = 0..8191]

Use NcML and file NetCDF to work with HDF5 handler



AVHRR¹ via Hyrax directly

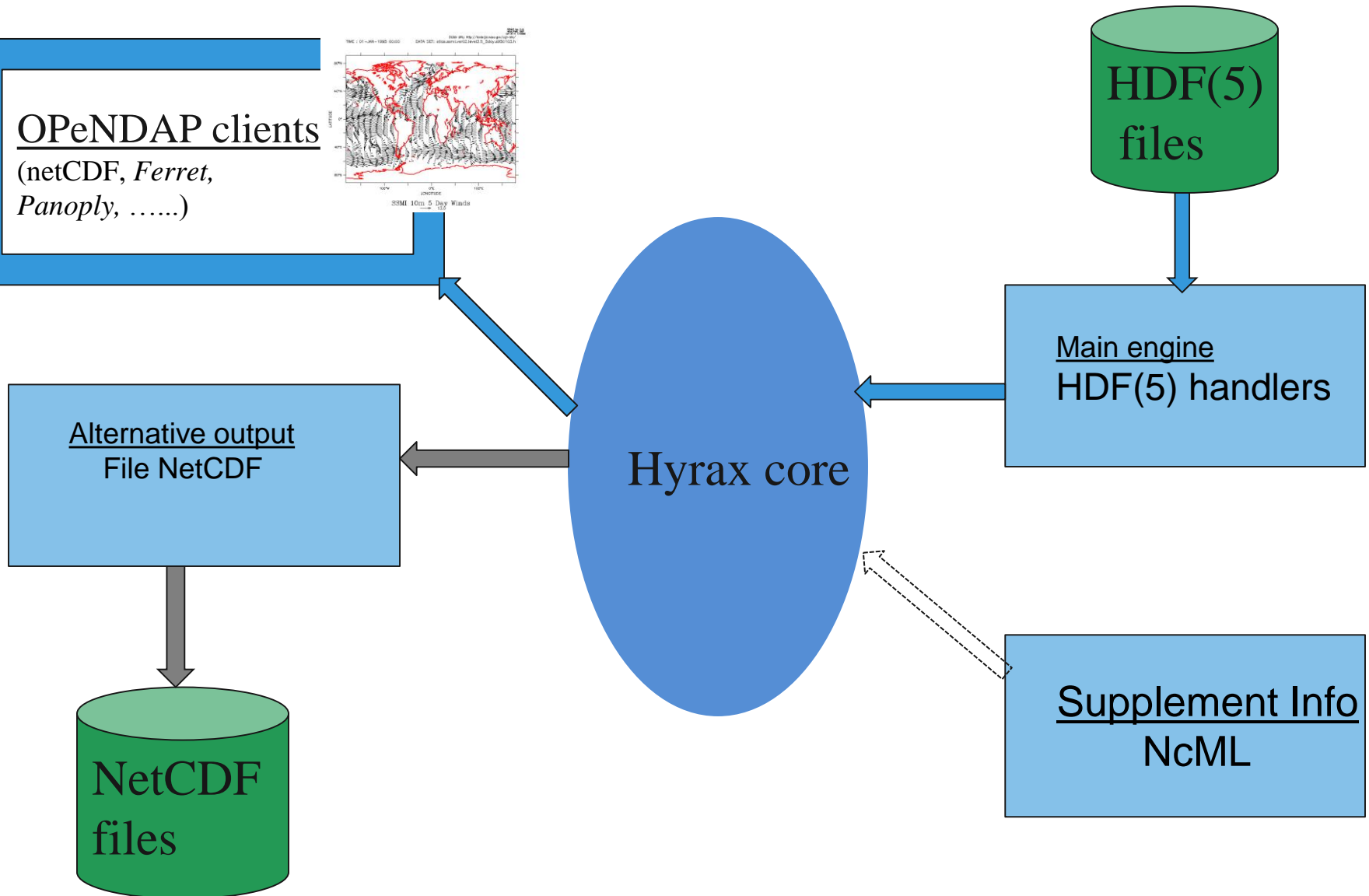
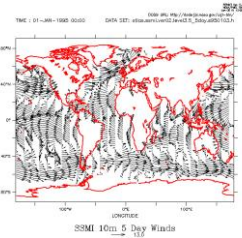


The NetCDF file of AVHRR¹
via Hyrax file NetCDF module

Service Chain to access HDF5 via Hyrax

OPeNDAP clients

(netCDF, Ferret,
Panoply,)



DAP¹4 support in the HDF5 handler

- CF option
 - DAP4 strictly mapped from DAP2
 - Dataset Metadata Response (DMR) replaces Dataset Descriptor Structure (DDS) and Dataset Attribute Structure (DAS)
- Default option
 - HDF5 group to DAP4 group
 - HDF5 signed 8-bit and 64-bit integers to DAP4
 - HDF5 dimensions following the NetCDF-4 to DAP4 mapping
 - Fully support NetCDF-4 (both classic and enhanced)

Interoperability enhancement

- CF option with DAP2
 - Provide a way for service providers to check if there are any objects ignored when mapping from HDF5 to DAP2
 - An example: DAP2 doesn't support 64-bit integer, HDF5 supports

Performance Improvement

- Reducing DDS and DAS access time
 - Caching DDS and DAS in memory
 - Caching DAS in the disk
- Reducing data access time
 - Caching the raw in the disk
 - Best if the data is compressed

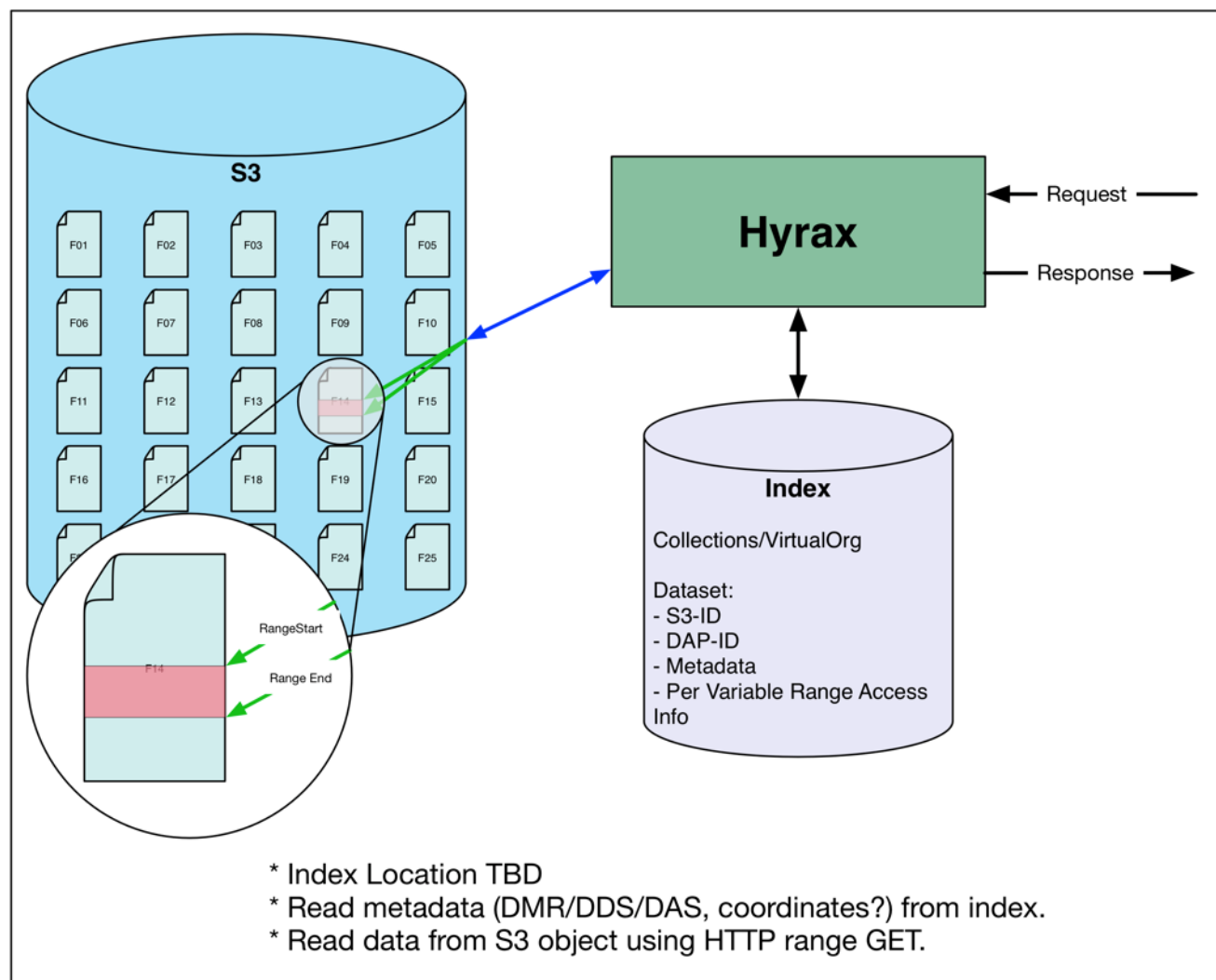
Other New Features

- Support the access of HDF-EOS5 sinusoidal projection in the HDF5 OPeNDAP handler

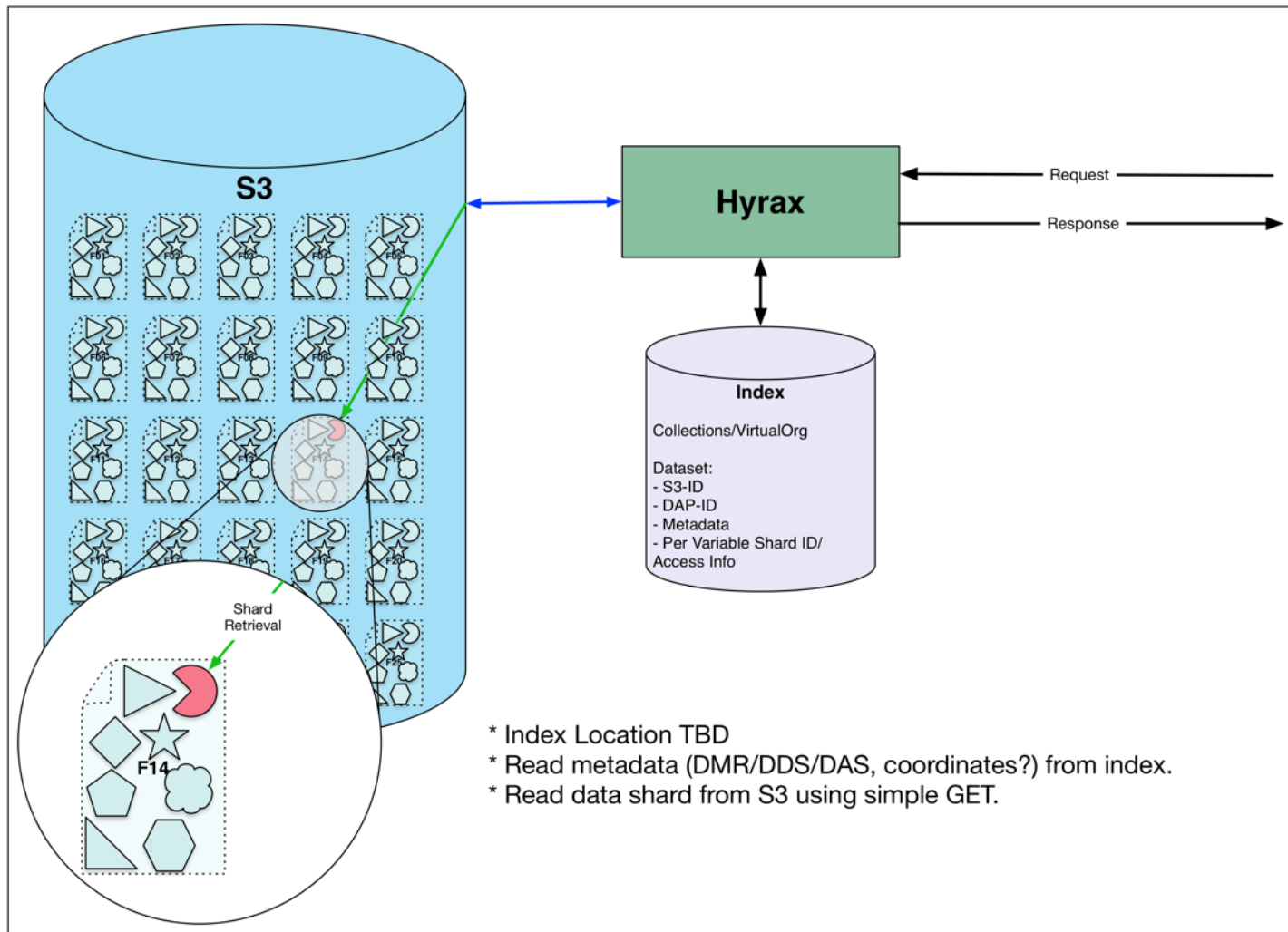
ACCESS HDF5 via Hyrax in Cloud

- Three architectures
- HDF5 handler can be enhanced for the future work of Architectures 2 and 3

Archit. #2: Files With HTTP¹ Range-Gets



Archit. #3: HDF5 Datasets as S3¹ Objects



What can be improved?

- Current
 - A2: Range-Gets index per chunk
 - A3: make an HDF5 chunk in a variable as an S3 object
- Future
 - Add an option to the HDF5 handler
 - Range-Gets index per variable
 - Make an HDF5 variable as an S3 object

This work was supported by
NASA/GSFC under Raytheon Co.
contract number NNG15HZ39C

Raytheon